



What is semiempirical molecular orbital theory approximating?

Johannes T. Margraf¹ · Pavlo O. Dral²

Received: 2 January 2019 / Accepted: 21 March 2019 / Published online: 16 April 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

We elucidate the approaches used to incorporate electron correlation in existing semiempirical molecular orbital theory (SEMO) methods and compare them with the techniques used in other quantum chemical methods. After analyzing expressions for electron correlation in *ab initio* wavefunction theory, density functional theory, and density functional-based tight-binding (TB) methods, we suggest a framework for developing hybrid TB-SEMO methods. We provide a numerical proof-of-concept for such a method based on the OM2 method.

Keywords Semiempirical molecular orbital theory · NDDO · Correlated orbital theory · Electron correlation · OM2 · MNDO

Introduction

Semiempirical molecular orbital theory (SEMO) is an umbrella term for a family of methods that originated in the work of John Pople in the 1960s as approximations to minimal basis Hartree–Fock (HF) theory [1–3]. The main hallmarks of these methods are the neglect of the atomic orbital (AO) overlap matrix ($S = \mathbf{1}$) in the secular equation and the neglect of a large number of (multi-center) two-electron integrals. While these approximations were originally intended to simply maintain the invariance properties of HF, they could partially be justified later on [4–6].

Even though it quickly became apparent that *ab initio* minimal basis HF is a woefully inadequate model chemistry, SEMO methods have survived to this day [7–13]. This is due to a shift in philosophy most prominently associated with Michael Dewar and Michael Zerner [14–18]. Instead of understanding SEMO as an approximation to HF, they simply used it as a parametric framework, which was fitted to experimental data. This allowed them to obtain thermochemical

(Dewar) and spectroscopic (Zerner) predictions with (at the time) unrivaled accuracy and efficiency. Importantly, the physical basis of the SEMO equations ensured that these methods were reasonably transferable, compared to fully empirical force-fields. At the same time, two different approaches to the development of the SEMO methods emerged: the first approach is exploiting parametric flexibility in a “brute-force” fitting strategy on as large and diverse training set as possible and the second approach is pursuing improvement of the physical model to be parametrized on a carefully selected limited training set [19–21].

The fully empirical route of the first approach brings the danger of unexpected (and sometimes catastrophic) failures for systems outside the training set. As an example, to improve hydrogen bond energies, additional Gaussian potentials were introduced in AM1 (and used in most PM_x methods) [13, 22, 23]. These corrections do improve hydrogen bond geometries (though not necessarily energies [24]), but they lead to non-covalent alkyl-alkyl equilibrium distances of below 1 Å. Ironically, dispersion corrections to PM6 therefore contain an additional empirical repulsive potential between hydrogen atoms [25, 26]. While this solves the specific issue of alkyl-alkyl interactions, it will likely cause new problems elsewhere (e.g., in the covalent bond of H₂). Overall, patching problems caused by empirical corrections with new empirical corrections resembles the old theory of epicycles in Ptolemaic astronomy.

In the second approach, the underlying non-parametric SEMO model is modified to come as close as possible to the parent minimal basis HF and to retain the speed of the SEMO methods, as in Thiel’s OM_x methods [6, 27]. Nevertheless,

This paper belongs to the Topical Collection Tim Clark 70th Birthday Festschrift.

✉ Johannes T. Margraf
johannes.margraf@ch.tum.de

¹ Chair for Theoretical Chemistry, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany

² Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany

recovering minimal basis HF is not the goal per se due to the known inadequacy of this method as a modern model chemistry. Instead, the hope is that given a better model (closer to HF), it should be easier to introduce and fit parameters to reproduce experimental and high-level theoretical properties and that the resulting method will be more robust in situations outside the training set. Indeed, the OMx methods greatly outperform methods of the first approach for very challenging systems with peculiar electronic structure [27, 28]. We argue, however, that the next-generation SEMO models should target not HF, but quantum chemical methods that include correlation in their model. These SEMO models should also remain as fast as modern general-purpose SEMO methods.

To approach the design of such next-generation SEMO models, we have to reconsider what the SEMO model is supposed to be approximating. In particular, it has long been argued that the empirical scaling of two-electron integrals in methods like MNDO (modified neglect of diatomic overlap) can be justified as an incorporation of dynamical correlation [14, 29]. If that is the case, then it is potentially misleading to look to HF for guidance regarding the desired form and properties of SEMO Hamiltonians. After all, HF by definition contains no electron correlation [30].

In this paper, we consider the question posed in the title in light of these considerations. To this end, we briefly review the equations of MNDO and related methods as they are derived in the context of HF. We then consider how correlation is introduced in other HF-like theories based on ab initio wavefunction theory (WFT) and Kohn–Sham (KS) density functional theory (DFT). From this analysis, we argue that the introduction of correlation effects into the SEMO Fock matrix is inconsistent with the use of a HF-like total energy expression. Alternative expressions are discussed, and a consistent parameterization strategy is proposed. Finally, some preliminary numerical results are shown, indicating the feasibility of improved, consistent SEMO methods.

Theory

Electron correlation in semiempirical Hamiltonians

The central equation to be solved in HF and other mean-field electronic structure methods is a generalized matrix eigenvalue problem in a basis of non-orthogonal AO basis functions [31]:

$$FC = \varepsilon SC, \quad (1)$$

with the Fock matrix F , the molecular orbital (MO) coefficient matrix C , the AO overlap matrix S and the diagonal MO eigenvalue matrix ε . This is typically solved by orthogonalizing the Fock matrix so that [32]:

$${}^\lambda F = S^{-1/2} F S^{-1/2} \quad (2)$$

and

$${}^\lambda C = S^{1/2} C \quad (3)$$

This leads to the canonical eigenvalue problem:

$${}^\lambda F {}^\lambda C = \varepsilon {}^\lambda C, \quad (4)$$

which can be solved by diagonalization of ${}^\lambda F$. In MNDO and related methods, the approximation is made that ${}^\lambda F = F$, so that Eq. (4) can be solved directly. The matrix elements of F resemble the HF ones, i.e.:

$$F_{\mu\nu} = T_{\mu\nu} + V_{\mu\nu} + \sum_{\lambda\sigma}^{AO} P_{\lambda\sigma} [(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)]. \quad (5)$$

Here, $T_{\mu\nu}$ and $V_{\mu\nu}$ are the one-electron kinetic and potential energy contributions, $(\mu\nu|\lambda\sigma)$ are two-electron repulsion integrals and $P_{\lambda\sigma}$ are density matrix elements defined as:

$$P_{\lambda\sigma} = \sum_i^{occ} n_i C_{\lambda i} C_{\sigma i}, \quad (6)$$

with the summation going over all occupied MOs. Here n_i are orbital occupation numbers.

For convenience, the first two terms on the r.h.s of Eq. (5) can be combined into the core Hamiltonian matrix H^{core} , whereas the remaining terms are grouped into the two-electron matrix G (i.e., $F = H^{core} + G$). The two-electron matrix can be further decomposed into the Coulomb (J) and exchange matrices (K).

In MNDO and related methods, $H_{\mu\nu}^{core}$ is directly given via empirical expressions. Meanwhile, G is strongly simplified via the neglect of diatomic differential overlap (NDDO) approximation, which leads to a neglect of most two-electron integrals according to [5]:

$$(\mu_A \nu_B | \lambda_C \sigma_D) = \delta_{AB} \delta_{CD} (\mu_A \nu_A | \lambda_C \sigma_C) \quad (7)$$

where the notation μ_A indicates that basis function μ is centered on atom A .

As an aside, this approximation (which amounts to a neglect of all three- and four-center integrals and some two-center two-electron integrals) is only partially justified in an orthogonal Löwdin basis for the valence-only minimal basis set, meaning that the assumption that ${}^\lambda F = F$ and the NDDO approximation are not entirely consistent with each other [6]. Consequently, the reintroduction of the generalized eigenvalue equation into one of the least accurate SEMO models led to noticeable improvements [33]. In addition, broadly used OMx methods remedy the imbalance in the one-electron integrals, which among the NDDO-retained integrals are most strongly affected by the orthogonalization [6, 7].

In the formulation of the MNDO method, Dewar and Thiel used an empirical expression for the remaining two-electron integrals ($\mu_A\nu_A|\lambda_C\sigma_C$). This expression is based on a multipole expansion of the densities $\mu_A\nu_A$ and $\lambda_C\sigma_C$, which is asymptotically correct for large distances r_{AC} . Meanwhile, in the limit $r_{AC}=0$, the values of the one-center two-electron integrals ($\mu_A\nu_A|\lambda_A\sigma_A$) should be recovered (which were derived from experimental data). In MNDO, this is achieved via the Dewar–Sabelli–Klopman (DSK) scaling function.

It was observed that this scaling leads to semi-empirical integrals that are smaller than the corresponding ab initio ones. Since the one-center integrals are derived from spectroscopic data (which obviously includes all electron correlation effects), this was interpreted as an average inclusion of dynamical correlation. Indeed, the mean-field potential of HF is known to be overly repulsive, an effect that is, e.g., reflected in erroneous one-electron properties such as dipole moments [34, 35].

To make this implicit correlation treatment more apparent, we can write the semi-empirical (SE) integrals as:

$$(\mu_A\nu_A|\lambda_C\sigma_C)_{SE} = (\mu_A\nu_A|\lambda_C\sigma_C) + V_{\mu\nu\lambda\sigma}^C(r_{AC}) \quad (8)$$

where we have introduced a correlation potential $V_{\mu\nu\lambda\sigma}^C(r_{AC})$, which is defined as the difference between the semiempirical and ab initio integrals. Using the definitions of the Coulomb and exchange matrix elements (Eq. 5), we can define an analogous matrix V^C for the correlation potential, which adds up the SE corrections (Eq. 8) to \mathbf{J} and \mathbf{K} . With this, we can reformulate the SEMO Fock matrix as:

$$F_{\mu\nu}^{SEMO} = H_{\mu\nu}^{core} + J_{\mu\nu} - K_{\mu\nu} + V_{\mu\nu}^C \quad (9)$$

It should be emphasized that the point of this reformulation of the MNDO equations is not to actually implement the method in this way, but rather to make the treatment of electron correlation more obvious, and to facilitate comparison with WFT and DFT methods. In general, Eq. (9) defines a correlated single determinant (independent particle) theory [36–38].

Correlation potentials in Kohn–Sham density functional theory

In an AO basis set, the KS equations closely resemble Eq. (9) [39]:

$$F_{\mu\nu}^{KS} = H_{\mu\nu}^{core} + J_{\mu\nu} + V_{\mu\nu}^{XC} \quad (10)$$

Here, $V_{\mu\nu}^{XC}$ is a matrix element of the exchange–correlation potential (V_{xc}) in the AO basis [40]. Indeed, hybrid functionals like B3LYP or PBE0 also include a scaled contribution of HF exchange ($K_{\mu\nu}$), further increasing the structural similarity between the KS and SEMO equations [41, 42].

Let us consider the properties of V_{xc} and the corresponding KS determinants. Most importantly, the exact V_{xc} is the potential that yields the correct ground-state density of the system [40, 43]. Consequently, a good approximation to V_{xc} should yield a determinant that accurately reflects the exact electron density and related one-electron properties such as dipole moments and molecular electrostatic potentials.

A second important property of the exact V_{xc} is that the eigenvalue of the highest occupied molecular orbital (HOMO) of the corresponding self-consistent KS determinant is exactly equal to the vertical ionization potential (IP) of the system [44]. Furthermore, Bartlett has formulated a more general IP theorem, based on the adiabatic time-dependent DFT equation [36, 45, 46]. According to this, all occupied orbital eigenvalues should correspond to vertical ionization potentials. Baerends and coworkers found this to be true for a broad range of molecules when using the statistical averaging of orbital potentials (SAOP) approximation to V_{xc} [47].

The exact V_{xc} is not known in general, but it can be determined for specific systems, e.g., via the optimized effective potential method or by inverting accurate electron densities [48–52]. Such studies offer numerical confirmation that the above conditions (in particular regarding ionization potentials) are indeed properties of the ideal V_{xc} . For general applications, many approximations to V_{xc} exist, which fulfill these conditions to a greater or lesser extent [45, 53].

In this context, it should be noted that the focus of most functional developers has been on the exchange–correlation functional (E_{xc} , see below), and not on the potential [54]. This is attractive because the corresponding potential can in principle be derived for any functional via the functional derivative with respect to the electron density ρ :

$$V_{xc}(\mathbf{r}) = \frac{\partial E_{xc}[\rho]}{\partial \rho(\mathbf{r})} \quad (11)$$

Meanwhile, the inverse is not true, i.e., it is in general not possible to derive a functional that corresponds to some known potential. However, the convenience of Eq. (11) is misleading. The consistent potentials derived in this way tend to be of poor quality, reflecting e.g., in an erroneous description of transition states and other self-interaction based problems. Burke and coworkers have characterized these problems as “density-driven” errors [55, 56]. It is telling that these issues can (to a large extent) be fixed by combining HF densities non-self-consistently with common density functionals [57, 58]. In other words, having no correlation potential seems to be better than having a poor one in many cases.

Correlation potentials in wavefunction theory methods

Conventionally, higher-level WFT methods are based on a canonical HF reference, without any correlation contribution to the Fock matrix. The HF equations yield the variationally optimal reference, i.e., the one with the lowest energy. There are, however, also alternative choices to be made, such as Brueckner, natural or orbital-optimized references [38, 59–63].

These alternative determinants by definition have a higher energy than the canonical HF one. At first sight, this may seem like a drawback, as it means that more energy has to be recovered by the correlation treatment. However, these references are optimal in different ways. For instance, the Brueckner determinant has the largest overlap with the exact wavefunction and orbital-optimization yields the determinant which minimizes the total energy (i.e., the Hartree–Fock energy, E_{HF} , plus the correlation energy, E_C) [64]. These determinants are therefore more directly tied to the correlated wavefunction. The electrons are not simply subject to the mean-field potential, but feel an additional correlation potential [38]. The corresponding generalized Fock matrices thus formally resemble Eq. (9).

As for DFT, we can ask what the properties of these correlation potential and the related determinants are. Hesselmann and Jansen showed that the one-electron and response properties of Brueckner determinant are in significantly better agreement with experiment than those of the uncorrelated HF determinant [35]. The same is true for natural orbitals, which are obtained from diagonalization of correlated one-electron reduced-density matrices. By construction, they therefore reflect the more accurate electron density of the correlated treatment. Similarly, orbital-optimized MP2 calculations have been shown to produce accurate spin densities [62]. Finally, there is a close relationship between the extended Brueckner Fock matrix and correlation corrections to Koopman's theorem, meaning that the correlation potential also improves the orbital energies [65, 66].

Though not usually counted as WFT methods, we should also mention electron propagator and Green's function approaches, such as the Outer Valence Green's Function and GW approximations [38, 65, 67–70]. Here, electrons interact with a screened potential that is described by a so-called “self-energy” term. Again, the methods can formally be written via a generalized Fock matrix resembling Eq. (9), with the self-energy corresponding to the correlation potential [65]. The eigenvalues of this equation then correspond to vertical IPs (and electron affinities).

Comparison between SEMO, DFT, and WFT methods

As a short summary of the previous three sections, we note that: 1) The effect of scaling two-electron integrals in SEMO corresponds to introducing a correlation potential into the Fock matrix. 2) The exact exchange–correlation potential in KS-DFT leads to a determinant that has the exact ground state electron density and occupied orbital eigenvalues, which correspond to vertical IPs. 3) Correlation potentials in WFT also improve the electron density and eigenvalue spectra (but not the energy of the determinant) relative to HF.

As both DFT and WFT are in principle exact, the corresponding exact correlation potentials must have the same properties. This should also be true for good approximations in both theories, although the KS potentials of popular density functional approximations like BLYP are actually rather poor [49]. While SEMO correlation potentials have not usually been considered explicitly, the parameterization of the methods typically uses reference dipole moments and (principal) IPs. The recent hpCADD-Hamiltonian of Clark and coworkers further includes molecular electrostatic potentials [71]. In light of the above discussion, this is a very good choice. Furthermore, the use of experimentally (or theoretically) derived one-center two-electron integrals ensures that the SEMO Hamiltonian is tied to the correlated ionization potentials of isolated atoms [72, 73].

In this sense, the established practice in the SEMO community is vindicated by our analysis. However, heats of formation and molecular geometries also play a very significant role in the parameterization. As we will discuss in the following section, this should be addressed independently.

Total energy expressions

The total electronic energy E_{el} in HF and SEMO is calculated via:

$$\begin{aligned} E_{el}^{HF} &= \frac{1}{2} \sum_{\mu\nu} P_{\mu\nu} \left[H_{\mu\nu}^{core} + F_{\mu\nu} \right] \\ &= \sum_{\mu\nu}^{AO} P_{\mu\nu} \left[H_{\mu\nu}^{core} + \frac{1}{2} G_{\mu\nu} \right] \end{aligned} \quad (12)$$

As is well known, this total electronic energy is not equal to the sum of occupied MO eigenvalues, sometimes referred to as the band energy E_{band} :

$$\begin{aligned} E_{band} &= \sum_i^{occ} n_i \varepsilon_i = \sum_i^{occ} \sum_{\mu\nu}^{AO} n_i C_{\mu i} C_{\nu i} F_{\mu\nu} \\ &= \sum_{\mu\nu}^{AO} P_{\mu\nu} \left[H_{\mu\nu}^{core} + G_{\mu\nu} \right] \end{aligned} \quad (13)$$

It follows that the band energy includes a double counting of two-electron terms, which can be removed to obtain the

total energy, in terms of E_{band} [74]:

$$E_{el}^{HF} = \sum_i^{occ} n_i \varepsilon_i - \frac{1}{2} \sum_{\mu\nu}^{AO} P_{\mu\nu} G_{\mu\nu} \quad (14)$$

This gives the total electronic energy of a single determinant in HF. However, as discussed above, the inclusion of a correlation potential does not improve (in fact worsens) the energy of the reference determinant. From a total energy perspective, the benefit of using a correlated determinant is only seen when it is combined with an explicit expression for the correlation energy, e.g., via coupled cluster (CC) or perturbation theory (so-called post-HF methods).

There has in fact been a long history of applying such post-HF treatments to SEMO Hamiltonians. Most prominently, the configuration interaction (CI) method can be combined with SEMO determinants to obtain excited state properties [8, 75–79]. While there is some concern with respect to double-counting between the implicit and explicit description of correlation, this does not seem to be a major issue as long as the CI treatment is mostly used to describe static correlation within a limited active space [78].

In contrast, explicit correlation treatments for ground-state properties are rarely used, although the MNDO/C method was explicitly parameterized for ground-state energetics, in combination with second-order perturbation theory [29]. This approach is formally well justified in our view, but the use of traditional correlated quantum chemistry methods in a minimal basis set does not yield satisfactory results, as high-angular momentum polarization functions are necessary to adequately describe the electron–electron cusp [80]. From a more practical point of view, using a WFT-based correlation energy expression significantly impacts the main advantage of SEMO methods, namely their computational efficiency.

It is therefore worthwhile to look at DFT, where the correlated total energy is obtained at a mean-field cost. This is achieved through the exchange–correlation functional E_{xc} . In analogy to Eq. (14), the total electronic energy in DFT can be written in terms of a sum of orbital energies and correction terms [74, 81]:

$$E_{el}^{DFT} = \sum_i^{occ} n_i \varepsilon_i - E_H[\rho] - \int V_{xc}(r) \rho(r) dr + E_{xc}[\rho] \quad (15)$$

Here, $E_H[\rho]$ is the Hartree energy (corresponding to $J_{\mu\nu}$ in Eq. 9), and the integral over V_{xc} corresponds to the contributions of $K_{\mu\nu}$ and $V_{\mu\nu}^C$. In other words, the first three terms in Eq. (15) are analogous to the semiempirical version of Eq. (14), where $V_{\mu\nu}^C$ is implicitly included in $G_{\mu\nu}$ via the scaled

integrals. However, Eqs. (12) and (14) contain no analogue of the last term $E_{xc}[\rho]$.

From the point-of-view advocated in this paper (namely that the semiempirical scaling of two-electron integrals corresponds to an implicit definition of a correlation potential analogous to V_{xc}), the use of Eqs. 12 or 14 to determine the electronic energy in SEMO is therefore inconsistent.

How can this inconsistency be resolved? The straightforward solution would be to include an explicit $E_{xc}[\rho]$. However, this is not ideal for several reasons. Firstly, the quality of SEMO electron densities is limited due to the minimal basis set and frozen core approximation. Secondly, $E_{xc}[\rho]$ is generally evaluated with numerical quadrature, which has negligible computational cost in the context of a full first-principles DFT calculation, but may significantly impact the efficiency of a SEMO approach. Finally, as we alluded to above, there is no simple prescription for obtaining the $E_{xc}[\rho]$, which corresponds to a given V_{xc} .

It is instead worth looking to semiempirical density functional based tight-binding (TB) schemes such as DFTB [82–86]. These methods also avoid the explicit evaluation of $E_{xc}[\rho]$, despite being approximations to DFT. In the simplest case, the total energy of a TB scheme is given by:

$$E_{tot}^{TB} = \sum_i^{occ} n_i \varepsilon_i + \frac{1}{2} \sum_{A,B}^{atoms} V_{rep}(r_{AB}) \quad (16)$$

Here, $V_{rep}(r_{AB})$ are pair-potentials, which fold in all effects from the last three terms of Eq. (15), as well as the core–core repulsion between atoms A and B. This term is purely empirical, but it has been shown that the sum of these contributions is indeed approximately pairwise and short-ranged in DFT [74]. Formally, it resembles empirical modifications of the core–core terms that are also known in SEMO methods based on Eq. (12) (used, e.g., to improve the description of hydrogen bonds) [87]. Similar pairwise expressions have also been used by Grimme et al. to correct basis-set insufficiencies [88, 89].

Determining the total energy of a semiempirical Hamiltonian via Eq. (16) defines a hybrid TB-SEMO method. In the tight-binding literature, the short-ranged, repulsive nature of $V_{rep}(r_{AB})$ is attributed to a cancellation between double-counting terms, $E_{xc}[\rho]$ and the core–core repulsion [74]. This indicates that the use of Eq. (16) might be advantageous over two other alternatives, which could be used.

One of them is to add an additional nucleus–electron potential to the core–Hamiltonian [90]. This technique allows for description of noncovalent interactions, which are not described adequately by HF and the SEMO

methods without dispersion corrections, but are described adequately with post-HF methods including correlation explicitly. This recovers the long-range correlation contribution, but not the complete correlation energy.

Another approach is to model the effects of the correlation functional explicitly:

$$E_{el}^{SEMO} = \sum_i^{occ} n_i \varepsilon_i - \frac{1}{2} \sum_{\mu\nu}^{AO} P_{\mu\nu} G_{\mu\nu} + E_C^{SE} \quad (17)$$

Here, E_C^{SE} represents some semiempirical expression for the correlation energy, which remains to be defined, although empirical and machine-learning (ML)-based approximations for the correlation energy have been reported [91–94]. From a formal perspective, Eq. (17) mirrors Møller–Plesset perturbation theory. Specifically, if the Hamiltonian is partitioned as

$$\hat{H} = \hat{F} + \lambda \hat{V}, \quad (18)$$

the first and second terms on the r.h.s of Eq. (17) correspond to the MP0 and MP1 energies, whereas the last term in principle contains all higher orders (and in practice would be truncated, e.g., to second order). As mentioned above, the explicit calculation of the correlation energy in a minimal basis set is not desirable, however.

Independent of the final form of the total energy expression, the above arguments have important consequences with respect to the parameterization procedure. In the spirit of a correlated orbital theory, all SEMO parameters that directly affect the Fock matrix elements (electronic parameters) should be optimized to fulfill the known properties of a correlation potential. Specifically, only IPs and one-electron properties (such as dipole moments) should be used as reference data at this stage. Clearly, this will lead to a poorer performance for thermochemistry and molecular geometries, if the electronic energy is determined according to Eqs. (12) or (14). These errors should then be corrected by a further parametric expression (e.g., $V_{rep}(r_{AB})$), which now includes the effects of $E_{xc}[\rho]$.

Importantly, the parameterization of this term is not performed simultaneously with the electronic parameters, but subsequently. In other words, first the electronic parameters are optimized to correctly describe a reference set of IPs and dipole moments (for molecules with fixed geometries). Second, the pair potentials are parameterized to correctly describe thermochemistry and geometries. Besides being physically motivated, this separation of potential and energy functional also has practical advantages for the parameterization process. We expect that the error function for both steps should be significantly smoother, compared to when thermochemical and electronic reference data is included simultaneously.

Results and discussion

In the previous section, we analyzed the nature of the implicit description of electron correlation in SEMO methods. This analysis showed that (1) a single-particle correlation potential leads to improved one-electron properties (orbital energies, dipole moments, etc.) and (2) evaluating the total energy of a correlated determinant with the HF energy expression does not lead to improved energies. We concluded that the parameterization of SEMO Hamiltonians should focus on one-electron properties (rather than energies and molecular geometries), and that the SEMO total energy expression needs an additional term to account for the correlation energy. Unfortunately, the full development and parameterization of a new SEMO method along these lines is beyond the scope of this manuscript. In the following, we provide some numerical evidence that such a method is feasible, however.

Specifically, we construct a prototype method using the unmodified SEMO formalism for the Fock matrix. We decided to build this TB-SEMO method based on the OM2 [7, 95, 96] Hamiltonian, which is known [27, 97] to be among the most robust NDDO-based SEMO methods. The Hamiltonian is reparameterized using only one-electron properties as reference data. We then apply this Hamiltonian in an energy expression like Eq. (16), and explore the properties of the corresponding repulsive potential $V_{rep}(r_{AB})$. All OM2 calculations were performed with the development version of the MNDO program [98].

We started construction of the TB-OM2 method by reparameterization of the OM2 technique to minimize the sum of squares of errors (SSQ) in IPs (calculated according to Koopmans' theorem) and dipole moments for the set of fixed reference geometries. The training set was the CHNO set [27] used in parametrization of the OM2 method. We optimized all OM2 parameters except for the one-center two-electron integrals and parameters for the effective core potentials. First, we optimized each parameter one after the other. Second, we simultaneously re-optimized all parameters obtained in the first step. We used the Subplex optimization algorithm as implemented in the NLOpt library [99, 100]. Weighting factors for errors in IPs and dipole moments were adjusted so that their SSQs were numerically equal for calculations with the standard OM2 parameters used as the initial guess in reparameterization ($\frac{w_{IP}}{w_{Dipole}} = 1.04259$). Optimized parameter values of TB-OM2 are given in Table 1.

The thus obtained TB-OM2 Hamiltonian performs systematically much better for IPs and dipole moments than the standard OM2 Hamiltonian as is clear from the correlation plots between calculated and reference values for the CHNO set (Figs. 1 and 2). The correlation coefficients are higher, the slope of the linear trend line is closer to unity, and its y-intercept is closer to zero for TB-OM2 compared to OM2.

Table 1 Parameters of the TB-OM2 method and their deviations from the standard OM2 parameters (in parentheses) for H, C, N, and O elements. See Ref. [7] for the definitions of the OM2 parameters

	H	C	N	O
U_{ss} (eV)	-12.93865065 (2.3%)	-54.19555385 (4.9%)	-72.15968201 (-3.0%)	-122.69410683 (20.5%)
U_{pp} (eV)		-39.76564918 (0.1%)	-57.38765195 (-0.4%)	-78.64608937 (-0.4%)
ζ (au)	1.48992759 (1.1%)	1.40107513 (-1.4%)	1.63045334 (22.4%)	1.21940062 (-21.4%)
β_s (eV bohr ^{-1/2})	-3.76196154 (10.0%)	-7.05958647 (-2.1%)	-11.17779312 (3.1%)	-13.34246104 (25.3%)
β_p (eV bohr ^{-1/2})		-5.13983530 (24.0%)	-11.93774750 (56.6%)	-9.19185602 (6.4%)
β_π (eV bohr ^{-1/2})		-5.66899547 (-5.1%)	-9.68090088 (4.3%)	-13.42610721 (45.7%)
$\beta_s(X-H)$ (eV bohr ^{-1/2})		-6.35752364 (0.9%)	-18.98337612 (99.9%)	-5.51768151 (-15.7%)
$\beta_p(X-H)$ (eV bohr ^{-1/2})		-4.25869645 (5.3%)	-10.92895254 (28.4%)	-10.17896950 (0.7%)
α_s (au)	0.05484270 (-17.0%)	0.08562385 (-5.3%)	0.08448189 (-5.9%)	0.11299661 (-13.5%)
α_p (au)		0.05640475 (3.5%)	0.09090711 (3.8%)	0.07827396 (-18.7%)
α_π (au)		0.10045879 (-1.6%)	0.13311154 (1.1%)	0.17413733 (33.2%)
$\alpha_s(X-H)$ (au)		0.10425187 (7.8%)	0.11930093 (4.4%)	0.26066376 (134.6%)
$\alpha_p(X-H)$ (au)		0.06717192 (27.1%)	0.08192943 (-23.2%)	0.07964335 (-33.0%)
F_1	0.24980893 (-15.5%)	0.35381926 (-29.2%)	0.48099152 (-24.9%)	1.09527557 (-13.4%)
F_2	1.42645320 (1.8%)	0.55289933 (-23.5%)	0.00195808 (-99.0%)	2.04113892 (77.7%)
G_1	0.76452681 (17.1%)	0.12392396 (-41.8%)	0.02646063 (-81.0%)	0.49849623 (76.1%)
G_2	0.60740044 (-33.1%)	1.02682114 (3.5%)	3.04394865 (260.8%)	0.07074015 (-91.0%)

The mean absolute errors (MAEs) in IPs are 0.13 and 0.30 eV at TB-OM2 and OM2, respectively. MAEs in dipole moments are 0.13 and 0.27 D at TB-OM2 and OM2, respectively. We note that the errors of OM2 somewhat decrease for IPs (MAE of 0.26 eV) and dipole moments (MAE of 0.25 D) when full optimizations are performed at this level of theory (this is how the OM2 method was parametrized), but they remain much higher than errors of TB-OM2 [27]. The linear trend lines are very similar for IPs and dipole moments calculated with OM2 regardless of whether geometry optimizations were performed or not at this level of theory.

The TB-OM2 Hamiltonian thus performs quite well for the targeted properties. It is, of course, not very surprising that the new method outperforms the standard OM2 for these properties, given the parameterization. However, we do find it

notable that such high accuracy can be achieved in absolute terms. This means that the minimal-basis NDDO framework can faithfully represent the exact one-electron potentials. This confirms that SEMO methods are an attractive basis for the development of an efficient correlated orbital theory.

Nevertheless, the new parameters are not suitable for total energy calculations following the conventional theory (i.e., based on Eq. 12). Properties depending on total energies are described poorly, e.g., errors in heats of formation at 298 K exceed 100 kcal/mol for most of the molecules in the CHNO set. Errors in relative energies are also unacceptably large. Geometry optimizations lead to wrong structures.

These deficiencies can potentially be corrected with a new energy expression following Eq. (16). While developing these new approximations is beyond the scope of this study, it is

Fig. 1 Correlation between reference ionization potentials (IP_{ref}) and IPs calculated with the OM2 (left) and TB-OM2 (right) methods on the reference geometries (IP_{calc})

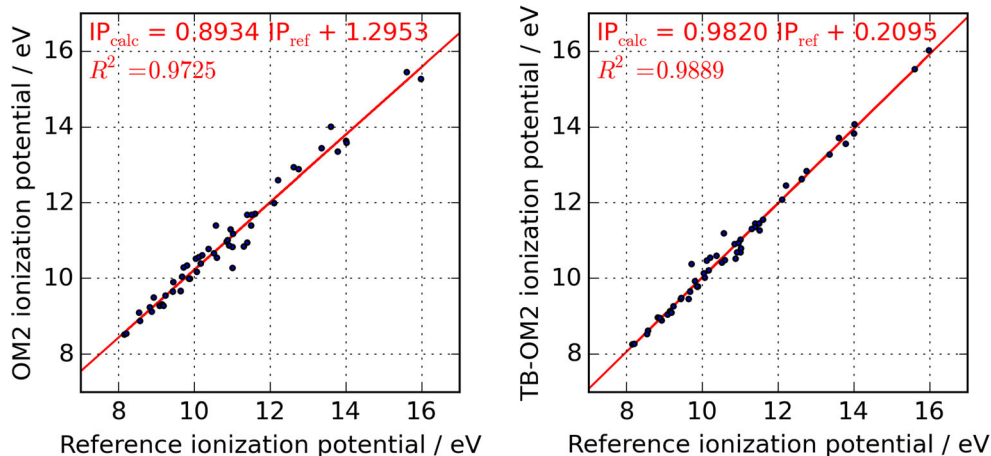
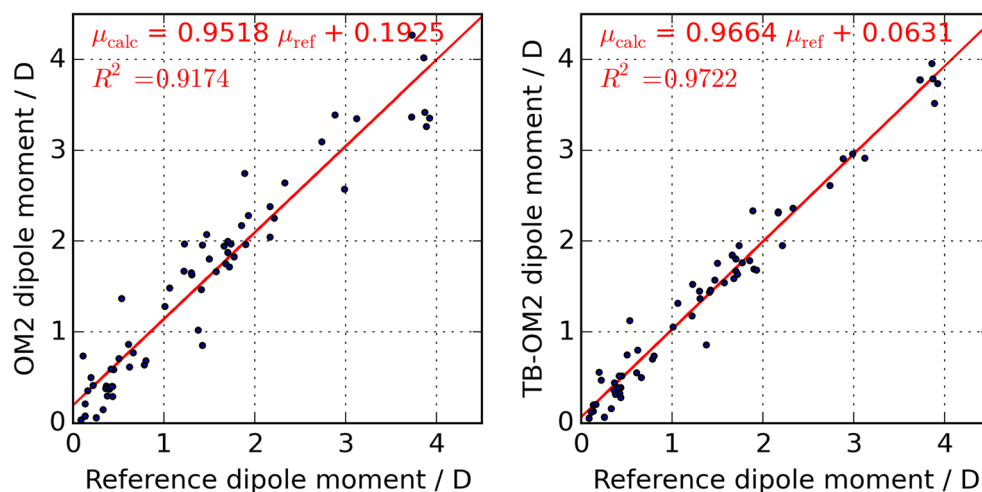


Fig. 2 Correlation between reference dipole moments (μ_{ref}) and dipole moments calculated with the OM2 (left) and TB-OM2 (right) methods on the reference geometries (μ_{calc})



instructive to investigate the properties of E_{band} (i.e., the sum of occupied orbital energies). This energy is the first term in Eq. (16) (and 17) and should recover most of the long-ranged electronic interactions. We can therefore use E_{band} to probe the properties of the unknown term in Eq. (16), i.e., $V_{\text{rep}}(r_{AB})$.

To this end, we apply a bond-projection procedure used in the DFTB literature [101]. Assuming that $V_{\text{rep}}(r_{AB})$ is a short-ranged pairwise potential, its values can be determined from rigid potential energy surface scans along bond coordinates. This was performed for the symmetric stretch of all C–H bonds in methane and along the C–C bond in ethane. Figures 3 and 4 show the forces obtained by differentiation of the PES with respect to these coordinates. As a reference, the same curves were computed at the PBE0/def2-TZVP level.

In both cases, the curves obtained from E_{band} are in reasonable agreement with the reference at large distances (i.e., $r > 2.5$ Å for C–C and $r > 1.5$ Å for C–H). This is because the systems dissociate into neutral atoms, meaning that long-range core–core interaction are screened. Furthermore, electronic effects are well

represented by E_{band} in this regime. Meanwhile, the curves deviate strongly in the bonding and repulsive regions, where this is not the case. The plots also show the difference between both curves, ΔF . The pairwise potential correcting the error of E_{band} (i.e., $V_{\text{rep}}(r_{AB})$) can now be obtained by integrating ΔF (see Fig. 5).

The resulting potentials behave analogously to what is typically observed in DFTB, i.e., they are short-ranged and mostly repulsive. Importantly, they tend to zero between the typical first and second nearest neighbor distances. This allows choosing appropriate cutoff values for these potentials, an essential requirement for a robust pairwise correction.

While these initial results are encouraging, the important question of how transferable the potentials are to other bond types and hybridizations remains open. Preliminary investigations for acetylene and ethylene were inconclusive because wavefunction instabilities were observed for the dissociation curves of these systems. Indeed, semiempirical methods are very prone to such instabilities, in particular for π -systems [102, 103].

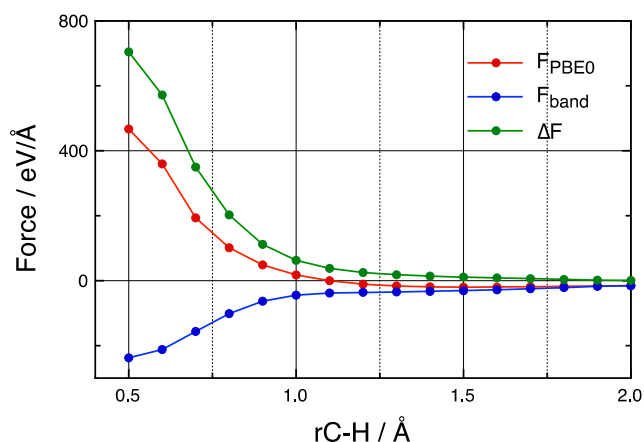


Fig. 3 Plot of forces obtained from the differentiation of a rigid PES scans along the C–H bond length in methane

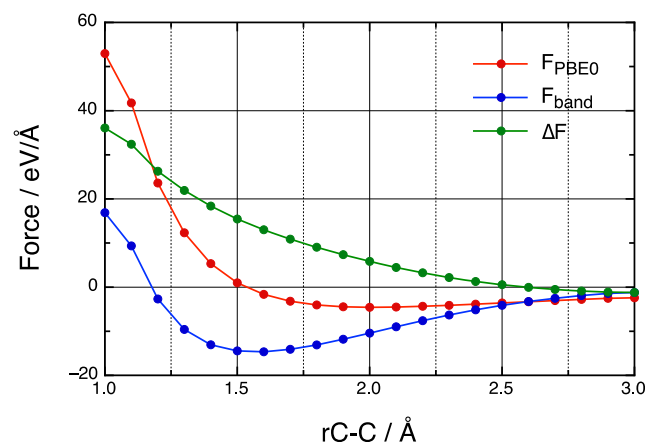


Fig. 4 Plot of forces obtained from the differentiation of a rigid PES scans along the C–C bond length in ethane

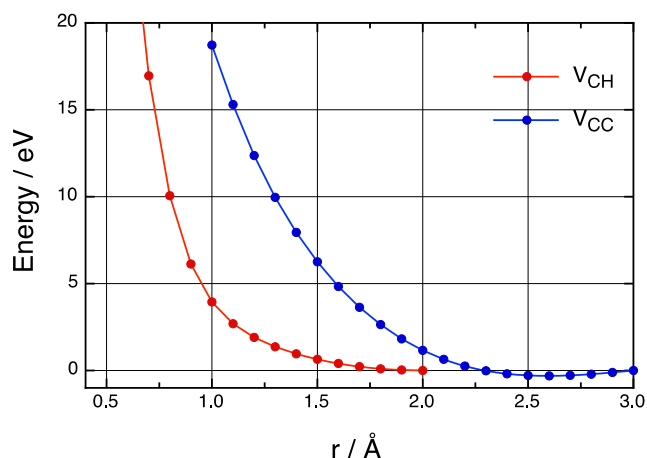


Fig. 5 Pairwise potentials obtained by integrating the ΔF curves in Figs. 3 and 4. In the case of methane, the potential is divided by the number of C–H bonds (four)

Consequently, the bond-projection scheme applied above is likely not well suited for determining a ‘general’ $V_{rep}(r_{AB})$. Instead, a force-matching approach for equilibrium structures may be a better choice.

Conclusions

In this work, we discussed the approaches used to incorporate electron correlation in the existing semiempirical molecular orbital (SEMO) theory, ab initio wavefunction theory, density functional theory, and density functional-based tight-binding (TB) methods. We outlined the design of next-generation TB-SEMO methods, which includes correlation via explicit DFTB-like pair-potentials (TB part of the model) also incorporating core–core repulsions. In these methods, the parametrization of the SEMO part of the model targets only ionization potentials and dipole moments and then it enters the TB-SEMO model via the sum of orbital energies. We provide a numerical proof-of-concept for such a method based on the OM2 method. The optimized parameters of this experimental TB-OM2 method provide a significant improvement over the OM2 method for the ionization energies and dipole moments. It was then demonstrated that the pair-wise potentials needed to complete the TB-OM2 model are short-ranged and mostly repulsive in analogy to the DFTB case.

Acknowledgements JTM thanks the Alexander-von-Humboldt Foundation and the Technical University of Munich for financial support. PD thanks the European Research Council for financial support through an ERC Advanced Grant (OMSQC). He also thanks Xin Wu for providing his parametrization program. We would like to dedicate this paper to Tim Clark who has been a mentor to both of us during our doctoral studies and ever since.

References

- Pople JA, Segal GA (1965) Approximate self-consistent molecular orbital theory. II. Calculations with complete neglect of differential overlap. J Chem Phys 43:S136–S151. <https://doi.org/10.1063/1.1701476>
- Pople JA, Santry DP, Segal GA (1965) Approximate self-consistent molecular orbital theory. I. Invariant procedures. J Chem Phys 43:S129–S135. <https://doi.org/10.1063/1.1701475>
- Pople JA (1953) Electron interaction in unsaturated hydrocarbons. Trans Faraday Soc 49:1375. <https://doi.org/10.1039/tf9534901375>
- Kolb M, Thiel W (1993) Beyond the MNDO model: methodical considerations and numerical results. J Comput Chem 14:775–789. <https://doi.org/10.1002/jcc.540140704>
- Chandler GS, Grader FE (1980) A re-examination of the justification of neglect of differential overlap approximations in terms of a power series expansion in S. Theor Chim Acta 54:131–144. <https://doi.org/10.1007/BF00554120>
- Wu X, Dral PO, Koslowski A, Thiel W (2019) Big data analysis of ab initio molecular integrals in the neglect of diatomic differential overlap approximation. J Comput Chem 40:638–649. <https://doi.org/10.1002/jcc.25748>
- Dral PO, Wu X, Spörkel L et al (2016) Semiempirical quantum-chemical orthogonalization-corrected methods: theory, implementation, and parameters. J Chem Theory Comput 12:1082–1096. <https://doi.org/10.1021/acs.jctc.5b01046>
- Dral PO, Clark T (2011) Semiempirical UNO-CAS and UNO-CI: method and applications in nanoelectronics. J Phys Chem A 115:11303–11312. <https://doi.org/10.1021/jp204939x>
- Thiel W (2014) Semiempirical quantum-chemical methods. Wiley Interdiscip Rev Comput Mol Sci 4:145–157. <https://doi.org/10.1002/wcms.1161>
- Margraf JT, Hennemann M, Meyer B, Clark T (2015) EMPIRE: a highly parallel semiempirical molecular orbital program: 2: periodic boundary conditions. J Mol Model 21:144. <https://doi.org/10.1007/s00894-015-2692-3>
- Hennemann M, Clark T (2014) EMPIRE: a highly parallel semiempirical molecular orbital program: 1: self-consistent field calculations. J Mol Model 20:2331. <https://doi.org/10.1007/s00894-014-2331-4>
- Ryan H, Carter M, Stenmark P et al (2016) A comparison of X-ray and calculated structures of the enzyme MTH1. J Mol Model 22:168. <https://doi.org/10.1007/s00894-016-3025-x>
- Stewart JJP (2013) Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J Mol Model 19:1–32. <https://doi.org/10.1007/s00894-012-1667-x>
- Dewar MJS, Thiel W (1977) Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen. J Am Chem Soc 99:4907–4917. <https://doi.org/10.1021/ja00457a005>
- Dewar MJS (1983) Development and status of MINDO/3 and MNDO. J Mol Struct 100:41–50. [https://doi.org/10.1016/0022-2860\(83\)90082-0](https://doi.org/10.1016/0022-2860(83)90082-0)
- Bingham RC, Dewar MJS, Lo DH (1975) Ground states of molecules. XXV. MINDO/3. Improved version of the MINDO semiempirical SCF-MO method. J Am Chem Soc 97:1285–1293. <https://doi.org/10.1021/ja00839a001>
- Dewar MJS, Lo DH (1972) Ground states of σ -bonded molecules. XVII. Fluorine compounds. J Am Chem Soc 94:5296–5303. <https://doi.org/10.1021/ja00770a026>
- Ridley J, Zerner M (1973) An intermediate neglect of differential overlap technique for spectroscopy: pyrrole and the azines. Theor Chim Acta 32:111–134. <https://doi.org/10.1007/BF00528484>

19. Dewar MJS, Healy EF, Holder AJ, Yuan Y-C (1990) Comments on a comparison of AM1 with the recently developed PM3 method. *J Comput Chem* 11:541–542. <https://doi.org/10.1002/jcc.540110413>
20. Stewart JJP (1990) Reply to “comments on a comparison of AM1 with the recently developed PM3 method”. *J Comput Chem* 11: 543–544. <https://doi.org/10.1002/jcc.540110414>
21. Clark T, Stewart JJP (2011) MNDO-like semiempirical molecular orbital theory and its application to large systems. *Computational methods for large systems*. Wiley, Hoboken, pp 259–286
22. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J Am Chem Soc* 107:3902–3909. <https://doi.org/10.1021/ja00299a024>
23. Stewart JJP (2004) Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1 and PM3 to more main group elements. *J Mol Model* 10:155–164. <https://doi.org/10.1007/s00894-004-0183-z>
24. Winget P, Selçuki C, Horn AHC et al (2003) Towards a “next generation” neglect of diatomic differential overlap based semiempirical molecular orbital technique. *Theor Chem Accounts* 110: 254–266. <https://doi.org/10.1007/s00214-003-0454-2>
25. Řezáč J, Hobza P (2012) Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J Chem Theory Comput* 8:141–151. <https://doi.org/10.1021/ct200751e>
26. Řezáč J, Hobza P (2011) A halogen-bonding correction for the semiempirical PM6 method. *Chem Phys Lett* 506:286–289. <https://doi.org/10.1016/j.cplett.2011.03.009>
27. Dral PO, Wu X, Spörkel L et al (2016) Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks for ground-state properties. *J Chem Theory Comput* 12:1097–1120. <https://doi.org/10.1021/acs.jctc.5b01047>
28. Korth M, Thiel W (2011) Benchmarking semiempirical methods for thermochemistry, kinetics, and noncovalent interactions: OMx methods are almost as accurate and robust as DFT-GGA methods for organic molecules. *J Chem Theory Comput* 7:2929–2936. <https://doi.org/10.1021/ct200434a>
29. Thiel W (1981) The MNDOC method, a correlated version of the MNDO model. *J Am Chem Soc* 103:1413–1420. <https://doi.org/10.1021/ja00396a021>
30. Löwdin P-O (1955) Quantum theory of many-particle systems. II. Study of the ordinary Hartree–Fock approximation. *Phys Rev* 97: 1490–1508. <https://doi.org/10.1103/PhysRev.97.1490>
31. Slater JC (1951) A simplification of the Hartree–Fock method. *Phys Rev* 81:385–390. <https://doi.org/10.1103/PhysRev.81.385>
32. Löwdin P-O (1950) On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *J Chem Phys* 18:365–375. <https://doi.org/10.1063/1.1747632>
33. Sattelmeyer KW, Tubert-Brohman I, Jorgensen WL (2006) NO-MNDO: reintroduction of the overlap matrix into MNDO. *J Chem Theory Comput* 2:413–419. <https://doi.org/10.1021/ct050174c>
34. Bak KL, Gauss J, Helgaker T et al (2000) The accuracy of molecular dipole moments in standard electronic structure calculations. *Chem Phys Lett* 319:563–568. [https://doi.org/10.1016/S0009-2614\(00\)00198-6](https://doi.org/10.1016/S0009-2614(00)00198-6)
35. Hesselmann A, Jansen G (1999) Molecular properties from coupled-cluster Brueckner orbitals. *Chem Phys Lett* 315:248–256. [https://doi.org/10.1016/S0009-2614\(99\)01251-8](https://doi.org/10.1016/S0009-2614(99)01251-8)
36. Bartlett RJ (2009) Towards an exact correlated orbital theory for electrons. *Chem Phys Lett* 484:1–9. <https://doi.org/10.1016/j.cplett.2009.10.053>
37. Beste A, Bartlett RJ (2004) Independent particle theory with electron correlation. *J Chem Phys* 120:8395–8404. <https://doi.org/10.1063/1.1691402>
38. Ortiz JV (2004) Brueckner orbitals, Dyson orbitals, and correlation potentials. *Int J Quantum Chem* 100:1131–1135. <https://doi.org/10.1002/qua.20204>
39. Pople JA, Gill PMW, Johnson BG (1992) Kohn–Sham density-functional theory within a finite basis set. *Chem Phys Lett* 199: 557–560. [https://doi.org/10.1016/0009-2614\(92\)85009-Y](https://doi.org/10.1016/0009-2614(92)85009-Y)
40. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. *Phys Rev* 140:A1133–A1138. <https://doi.org/10.1103/PhysRev.140.A1133>
41. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652. <https://doi.org/10.1063/1.464913>
42. Perdew JP, Ernzerhof M, Burke K (1996) Rationale for mixing exact exchange with density functional approximations. *J Chem Phys* 105:9982–9985. <https://doi.org/10.1063/1.472933>
43. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev* 136:B864–B871. <https://doi.org/10.1103/PhysRev.136.B864>
44. Katriel J, Davidson ER (1980) Asymptotic behavior of atomic and molecular wave functions. *Proc Natl Acad Sci* 77:4403–4406. <https://doi.org/10.1002/jpln.200900154>
45. Verma P, Bartlett RJ (2012) Increasing the applicability of density functional theory. III. Do consistent Kohn–Sham density functional methods exist? *J Chem Phys* 137:134102. <https://doi.org/10.1063/1.4755818>
46. Bartlett RJ, Ranasinghe DS (2017) The power of exact conditions in electronic structure theory. *Chem Phys Lett* 669:54–70. <https://doi.org/10.1016/j.cplett.2016.12.017>
47. Chong DP, Gritsenko OV, Baerends EJ (2002) Interpretation of the Kohn–Sham orbital energies as approximate vertical ionization potentials. *J Chem Phys* 116:1760–1772. <https://doi.org/10.1063/1.1430255>
48. Staroverov VN, Scuseria GE, Davidson ER (2006) Optimized effective potentials yielding Hartree–Fock energies and densities. *J Chem Phys* 124:141103. <https://doi.org/10.1063/1.2194546>
49. Bartlett RJ, Lotrich VF, Schweigert IV (2005) Ab initio density functional theory: the best of both worlds? *J Chem Phys* 123: 62205. <https://doi.org/10.1063/1.1904585>
50. Bartlett RJ, Grabowski I, Hirata S, Ivanov S (2005) The exchange-correlation potential in ab initio density functional theory. *J Chem Phys* 122:34104. <https://doi.org/10.1063/1.1809605>
51. Ryabinkin IG, Ospadov E, Staroverov VN (2017) Exact exchange-correlation potentials of singlet two-electron systems. *J Chem Phys* 147:164117. <https://doi.org/10.1063/1.5003825>
52. Gould T, Toulouse J (2014) Kohn–Sham potentials in exact density-functional theory at noninteger electron numbers. *Phys Rev A* 90:50502. <https://doi.org/10.1103/PhysRevA.90.050502>
53. Medvedev MG, Bushmarinov IS, Sun J et al (2017) Density functional theory is straying from the path toward the exact functional. *Science* 355:49–52. <https://doi.org/10.1126/science.aah5975>
54. Kohn W, Becke AD, Parr RG (1996) Density functional theory of electronic structure. *J Phys Chem* 100:12974–12980. <https://doi.org/10.1021/jp960669l>
55. Kim M-C, Sim E, Burke K (2013) Understanding and reducing errors in density functional calculations. *Phys Rev Lett* 111:73003. <https://doi.org/10.1103/PhysRevLett.111.073003>
56. Kim MC, Park H, Son S et al (2015) Improved DFT potential energy surfaces via improved densities. *J Phys Chem Lett* 6: 3802–3807. <https://doi.org/10.1021/acs.jpcllett.5b01724>
57. Verma P, Perera A, Bartlett RJ (2012) Increasing the applicability of DFT I: non-variational correlation corrections from Hartree–Fock DFT for predicting transition states. *Chem Phys Lett* 524: 10–15. <https://doi.org/10.1016/j.cplett.2011.12.017>

58. Oliphant N, Bartlett RJ (1994) A systematic comparison of molecular properties obtained using Hartree–Fock, a hybrid Hartree–Fock density-functional-theory, and coupled-cluster methods. *J Chem Phys* 100:6550–6561. <https://doi.org/10.1063/1.467064>
59. Lochan RC, Head-Gordon M (2007) Orbital-optimized opposite-spin scaled second-order correlation: an economical method to improve the description of open-shell molecules. *J Chem Phys* 126:164101. <https://doi.org/10.1063/1.2718952>
60. Reinhardt WP, Doll JD (1969) Direct calculation of natural orbitals by many-body perturbation theory: application to helium. *J Chem Phys* 50:2767. <https://doi.org/10.1063/1.1671446>
61. Bozkaya U, Sherrill CD (2013) Orbital-optimized coupled-electron pair theory and its analytic gradients: accurate equilibrium geometries, harmonic vibrational frequencies, and hydrogen transfer reactions. *J Chem Phys* 139:54104. <https://doi.org/10.1063/1.4816628>
62. Kossmann S, Neese F (2010) Correlated ab initio spin densities for larger molecules: orbital-optimized spin-component-scaled MP2 method. *J Phys Chem A* 114:11768–11781. <https://doi.org/10.1021/jp105647c>
63. Löwdin P-O (1955) Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. *Phys Rev* 97:1474–1489. <https://doi.org/10.1103/PhysRev.97.1474>
64. Chiles RA, Dykstra CE (1981) An electron pair operator approach to coupled cluster wave functions. Application to He₂, Be₂, and Mg₂ and comparison with CEPA methods. *J Chem Phys* 74:4544–4556. <https://doi.org/10.1063/1.441643>
65. Ortiz JV (2013) Electron propagator theory: an approach to prediction and interpretation in quantum chemistry. *Wiley Interdiscip Rev Comput Mol Sci* 3:123–142. <https://doi.org/10.1002/wcms.1116>
66. Cioslowski J, Piskorz P, Liu G (1997) Ionization potentials and electron affinities from the extended Koopmans' theorem applied to energy-derivative density matrices: the EKTMPn and EKTQCISD methods. *J Chem Phys* 107:6804–6811. <https://doi.org/10.1063/1.474921>
67. von Niessen W, Domcke W, Cederbaum LS, Kraemer WP (1977) Ionization potentials and vibrational structure in photoelectron spectra by a Green's function method: trans-HNNH, cis-HNNH, and 1,1-dihydrodiazine (H₂NN). *J Chem Phys* 67:44–51. <https://doi.org/10.1063/1.434539>
68. Van Setten MJ, Caruso F, Sharifzadeh S et al (2015) GW100: benchmarking G₀W₀ for molecular systems. *J Chem Theory Comput* 11:5665–5687. <https://doi.org/10.1021/acs.jctc.5b00453>
69. Hirata S, Hermes MR, Simons J, Ortiz JV (2015) General-order many-body greens function method. *J Chem Theory Comput* 11:1595–1606. <https://doi.org/10.1021/acs.jctc.5b00005>
70. Hirata S, Doran AE, Knowles PJ, Ortiz JV (2017) One-particle many-body Green's function theory: algebraic recursive definitions, linked-diagram theorem, irreducible-diagram theorem, and general-order algorithms. *J Chem Phys* 147:44108. <https://doi.org/10.1063/1.4994837>
71. Thomas HB, Hennemann M, Kibies P et al (2017) The hpCADD NDDO Hamiltonian: parametrization. *J Chem Inf Model* 57:1907–1922. <https://doi.org/10.1021/acs.jcim.7b00080>
72. Margraf JT, Claudino D, Bartlett RJ (2017) Determination of consistent semiempirical one-centre integrals based on coupled-cluster theory. *Mol Phys* 115:538–544. <https://doi.org/10.1080/00268976.2016.1200755>
73. Oleari L, Di Sipio L, De Michelis G (1966) The evaluation of the one-centre integrals in the semi-empirical molecular orbital theory. *Mol Phys* 10:97–109. <https://doi.org/10.1080/00268976600100161>
74. Foulkes WMC, Haydock R (1989) Tight-binding models and density-functional theory. *Phys Rev B* 39:12520–12536. <https://doi.org/10.1103/PhysRevB.39.12520>
75. Gadaczek I, Hintze KJ, Bredow T (2012) Periodic calculations of excited state properties for solids using a semiempirical approach. *Phys Chem Chem Phys* 14:741–750. <https://doi.org/10.1039/c1cp22871d>
76. Nelson T, Fernandez-Alberti S, Roitberg AE, Tretiak S (2014) Nonadiabatic excited-state molecular dynamics: modeling photophysics in organic conjugated materials. *Acc Chem Res* 47:1155–1164. <https://doi.org/10.1021/ar400263p>
77. Cremer D, Thiel W (1987) On the importance of size-consistency corrections in semiempirical MNDOC calculations. *J Comput Chem* 8:48–50. <https://doi.org/10.1002/jcc.540080106>
78. Tuna D, Lu Y, Koslowski A, Thiel W (2016) Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks of electronically excited states. *J Chem Theory Comput* 12:4400–4422. <https://doi.org/10.1021/acs.jctc.6b00403>
79. Fabiano E, Keal TW, Thiel W (2008) Implementation of surface hopping molecular dynamics using semiempirical methods. *Chem Phys* 349:334–347. <https://doi.org/10.1016/j.chemphys.2008.01.044>
80. Kong L, Bischoff FA, Valeev EF (2012) Explicitly correlated R12/F12 methods for electronic structure. *Chem Rev* 112:75–107. <https://doi.org/10.1021/cr200204r>
81. Harris J (1985) Simplified method for calculating the energy of weakly interacting fragments. *Phys Rev B* 31:1770–1779. <https://doi.org/10.1103/PhysRevB.31.1770>
82. Urban A, Reese M, Mrovec M et al (2011) Parameterization of tight-binding models from density functional theory calculations. *Phys Rev B* 84:155119. <https://doi.org/10.1103/PhysRevB.84.155119>
83. Margine ER, Kolmogorov AN, Reese M et al (2011) Development of orthogonal tight-binding models for Ti-C and Ti-N systems. *Phys Rev B* 84:155120. <https://doi.org/10.1103/PhysRevB.84.155120>
84. Gaus M, Goez A, Elstner M (2013) Parametrization and benchmark of DFTB3 for organic molecules. *J Chem Theory Comput* 9:338–354. <https://doi.org/10.1021/ct300849w>
85. Elstner M, Porezag D, Jungnickel G et al (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* 58:7260–7268. <https://doi.org/10.1103/PhysRevB.58.7260>
86. Seifert G (2007) Tight-binding density functional theory: an approximate Kohn-Sham DFT scheme. *J Phys Chem A* 111:5609–5613. <https://doi.org/10.1021/jp069056r>
87. Yilmazer ND, Korth M (2015) Enhanced semiempirical QM methods for biomolecular interactions. *Comput Struct Biotechnol J* 13:169–175. <https://doi.org/10.1016/j.csbj.2015.02.004>
88. Kruse H, Grimme S (2012) A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree–Fock and density functional theory calculations for large systems. *J Chem Phys* 136. <https://doi.org/10.1063/1.3700154>
89. Sure R, Grimme S (2013) Corrected small basis set Hartree–Fock method for large systems. *J Comput Chem* 34:1672–1685. <https://doi.org/10.1002/jcc.23317>
90. Kriebel M, Weber K, Clark T (2018) A Feynman dispersion correction: a proof of principle for MNDO. *J Mol Model* 24:338. <https://doi.org/10.1007/s00894-018-3874-6>
91. Welborn M, Cheng L, Miller TF (2018) Transferability in machine learning for electronic structure via the molecular orbital basis. *J Chem Theory Comput* 14:4772–4779. <https://doi.org/10.1021/acs.jctc.8b00636>

92. Margraf JT, Reuter K (2018) Making the coupled cluster correlation energy machine-learnable. *J Phys Chem A* 122:6343–6348. <https://doi.org/10.1021/acs.jpca.8b04455>
93. Ayala PY, Scuseria GE (2000) Atom pair partitioning of the correlation energy. *Chem Phys Lett* 322:213–218. [https://doi.org/10.1016/S0009-2614\(00\)00417-6](https://doi.org/10.1016/S0009-2614(00)00417-6)
94. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA (2015) Big data meets quantum chemistry approximations: the Δ -machine learning approach. *J Chem Theory Comput* 11:2087–2096. <https://doi.org/10.1021/acs.jctc.5b00099>
95. Weber W (2000) Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Implementierung, Parametrisierung und Anwendung. PhD thesis, Universität Zürich, Zürich, Hartung-Gorre Verlag
96. Weber W, Thiel W (2000) Orthogonalization corrections for semiempirical methods. *Theor Chem Accounts* 103:495–506. <https://doi.org/10.1007/s002149900083>
97. Silva-Junior MR, Thiel W (2010) Benchmark of electronically excited states for semiempirical methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J Chem Theory Comput* 6:1546–1564
98. Thiel W (2018) MNDO, development version. Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany
99. Rowan T (1990) Functional stability analysis of numerical algorithms. PhD Thesis, University of Texas at Austin, USA
100. Johnson S (2008) The NLOpt nonlinear-optimization package. <http://ab-initio.mit.edu/nlopt>
101. Koskinen P, Mäkinen V (2009) Density-functional tight-binding for beginners. *Comput Mater Sci* 47:237–253. <https://doi.org/10.1016/j.commatsci.2009.07.013>
102. Scuseria GE, Engelmann AR, Contreras RH (1982) Unrestricted Hartree-Fock instabilities in nuclear spin-spin coupling calculations. The MNDO method. *Theor Chim Acta* 61:49–57. <https://doi.org/10.1007/BF00573864>
103. Scuseria GE, Contreras RH (1980) Unrestricted Hartree-Fock instabilities in semiempirical CNDO/S and INDO/S calculations of spin-spin coupling constants. *Theor Chim Acta* 59:437–450. <https://doi.org/10.1007/BF00553399>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.